EUROSTEELMASTER2022

RINA

XIII Edition - Online course

European Advanced Training Course for the Worldwide Steel Sector

16 - 20 May 2022

With the support of

EUROFER
THE EUROPEAN STEEL ASSOCIATION

industriAll
EUROPEAN TRADE UNION

ESTEP

Federacciai

European Commission

Big Data Technologies in Downstream Steel Process

Raffaella Grieco
raffaella.grieco@rina.org

NewTech4Steel European RFCS project

CONTROL IN STEEL

RINA

Cloud and Big Data concepts (approaches, architectures and methods) applied in steel industry in particular at the fast-rolling processes (Lambda Architecture)
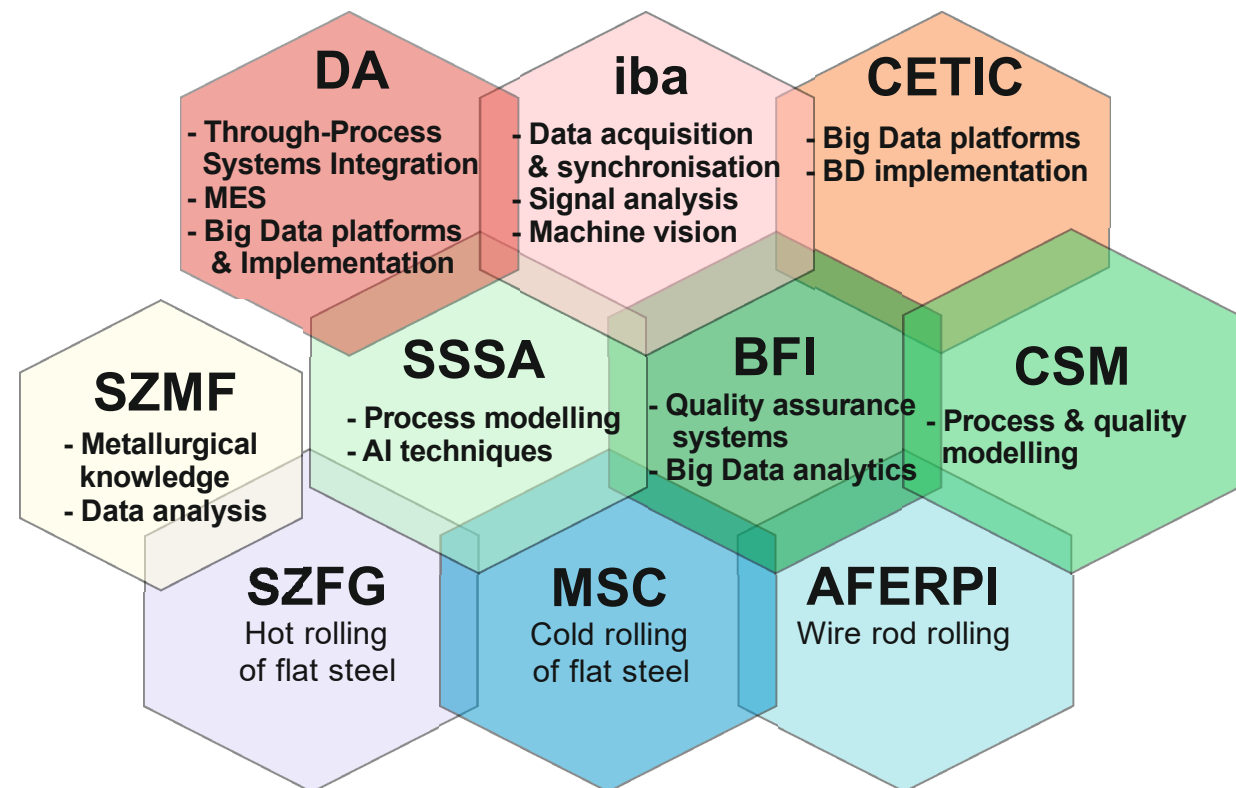
Project duration: 01.06.2018 – 31.05.2022

Providing and adapting infrastructures and architectures

Application of knowledge, methodologies and techniques to use cases

Industrial use cases
- **Hot-rolled wide strip**
- **Cold rolling area**
- **Wire rod rolling**

DA
- Through-Process Systems Integration
- MES
- Big Data platforms & Implementation

iba
- Data acquisition & synchronisation
- Signal analysis
- Machine vision

CETIC
- Big Data platforms
- BD implementation

SZMF
- Metallurgical knowledge
- Data analysis

SSSA
- Process modelling
- AI techniques

BFI
- Quality assurance systems
- Big Data analytics

CSM
- Process & quality modelling

SZFG
Hot rolling of flat steel

MSC
Cold rolling of flat steel

AFERPI
Wire rod rolling

Motivation

There is the demand to **increase the stability of production processes** and **the quality of the products** in the production of rolled steel.

- 1D/2D geometry (sensor, PLC) event – 10Hz
- Surface inspection (image)
- Process data (PLC, sensor, bus) event – 500Hz
- …
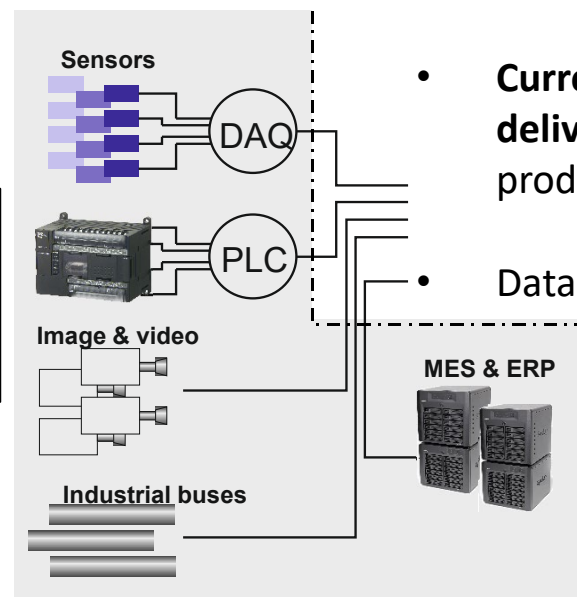
- **Current sensor equipment** at rolling facilities in steel production **delivers masses of data and information** about the process, the product and its quality

- Data are supplied **at high sample rates** and **high spatial resolutions**

Problem: process supervision **lacks** an **online or near-online exploitation** of the available data

Idea and Objective

Applicable methodologies to meet the requirements of online/near-online supervision of processes can be found in knowledge domains like **Data Analytics**, **Machine Vision** and in **Big Data techniques.**
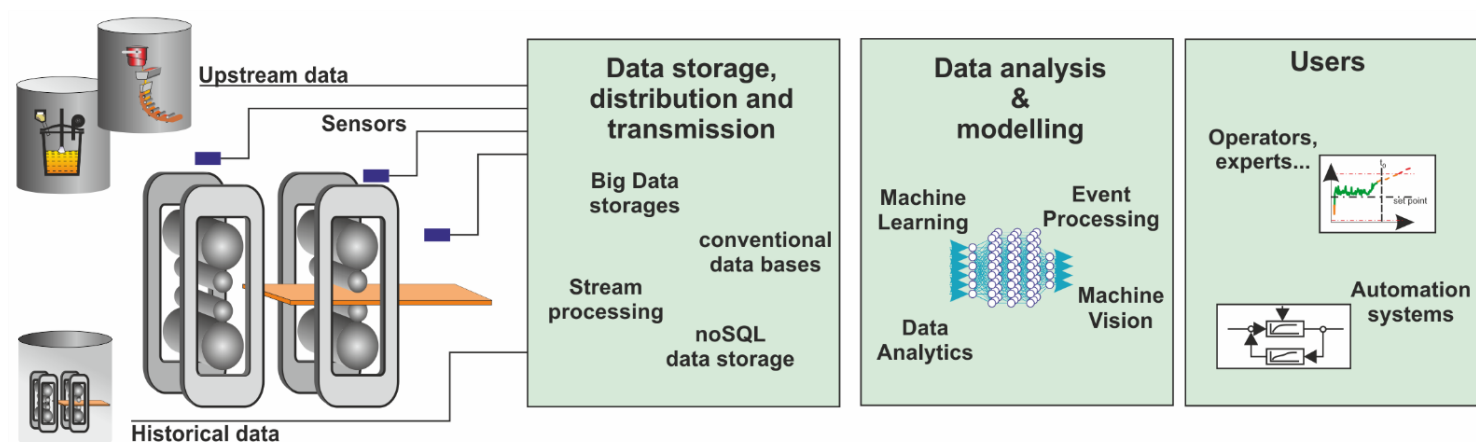

- **To toughen up the data storage systems** at the industrial sites by enhancing the existing solutions and integrating innovative tools from Big Data technology.

- **To establish a methodologies for data transmission** to achieve highest possible data throughput rates, necessary for the aimed online or near online processing.

- **To develop and apply newest technologies of Data Analytics** for analysis and modelling, because the existing data acquisition systems supply masses of data of various structures and contents, which have to be analysed and processed together.

Technical approach

The technical solutions based on
- the state-of-the-art equipment of the investigated production facilities (**brown-field implementation**),
- the experts' knowledge about the underlying processes and their physical relations (**model definition**),
- the application of innovative methodologies of modern data processing (**Big Data methodologies**).
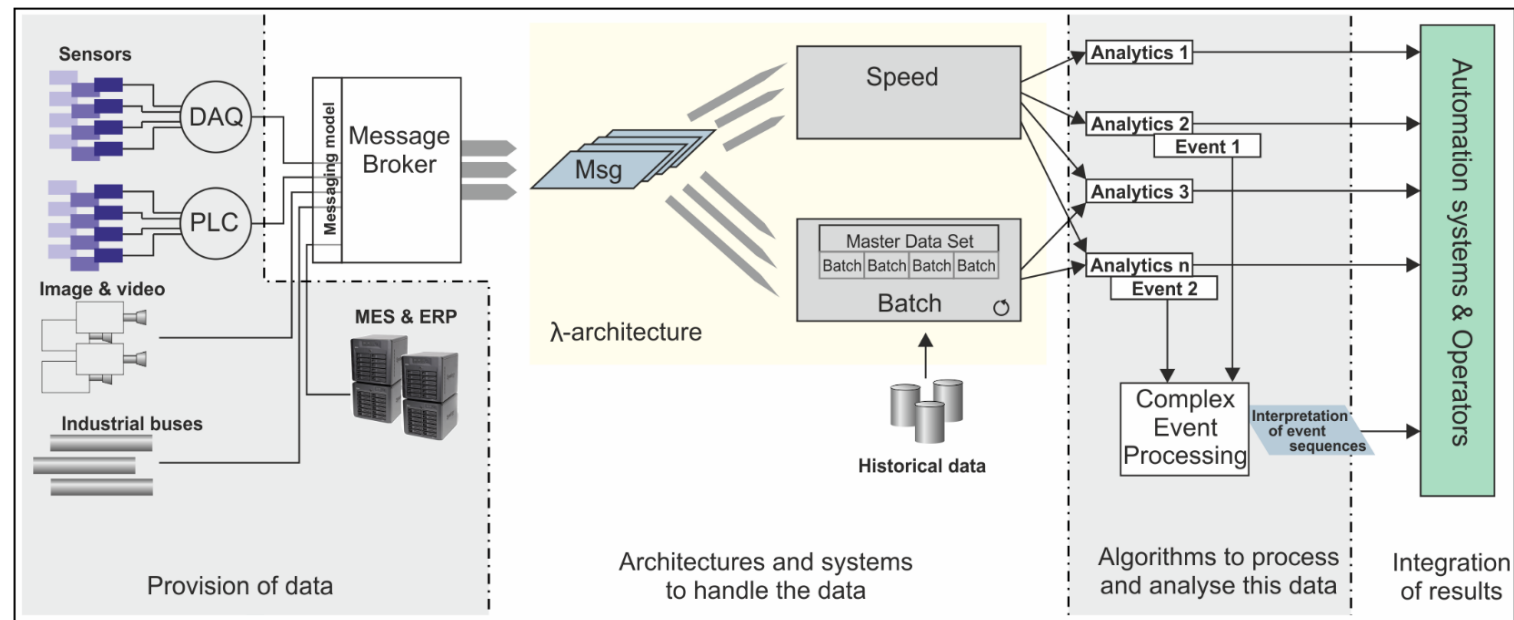
Lambda Architecture

The platform which merges the transmission of data for **online processing** (streaming) by a speed layer and for the **batch processing** where actual data are combined with those coming from historical data bases is the place where the Lambda architecture is applied.
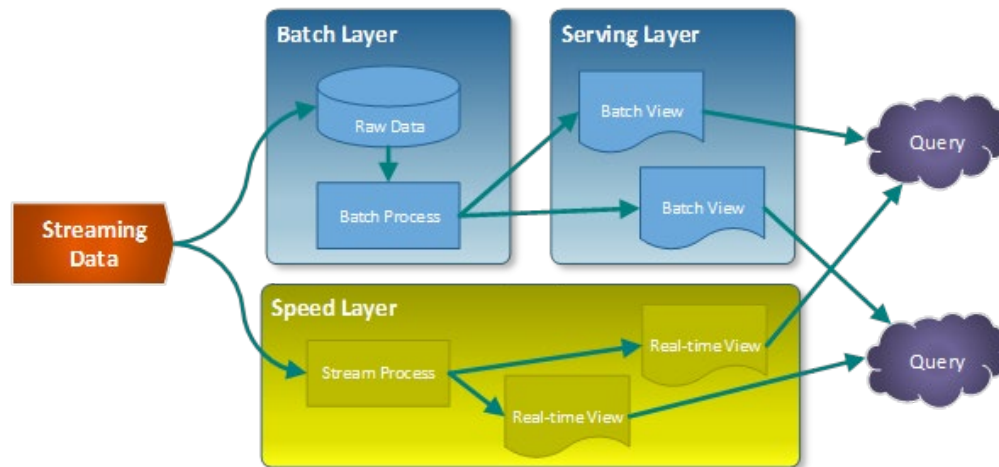
According to **Marz**, the essential requirements of the LA:
- Fault-tolerant and robust enough to withstand code failures and human errors
- The layers must be horizontally scalable.
- Low latency in order to achieve real-time results
- Easy to maintain and debug

Aim:

Development of a platform software solution based on Lambda Architecture to strengthen and optimize the processes of coils galvanization

a. **Forecast of top defects at HDG line due to flatness problem**
b. **Prediction of specific setup parameters to avoid or limit process problems and product defects during the galvanization**

Platform solution requirement:

- **Ingestion** and elaboration of high frequency variables (process parameters at Galvanizing line)
- **Storing** of length-based data (coil profile, mechanical charactetistics, stressometer matrix at CRM for each pass, Zinc profile top/bottom), time series and large data per coil (images) in datalake
- Heavy **data processing** for data preparation in the training cycle of ML
- Implementation, deployment and update of **ML models**
- Storage of results on relational-db
- HMI accessible via web

Plant Layout

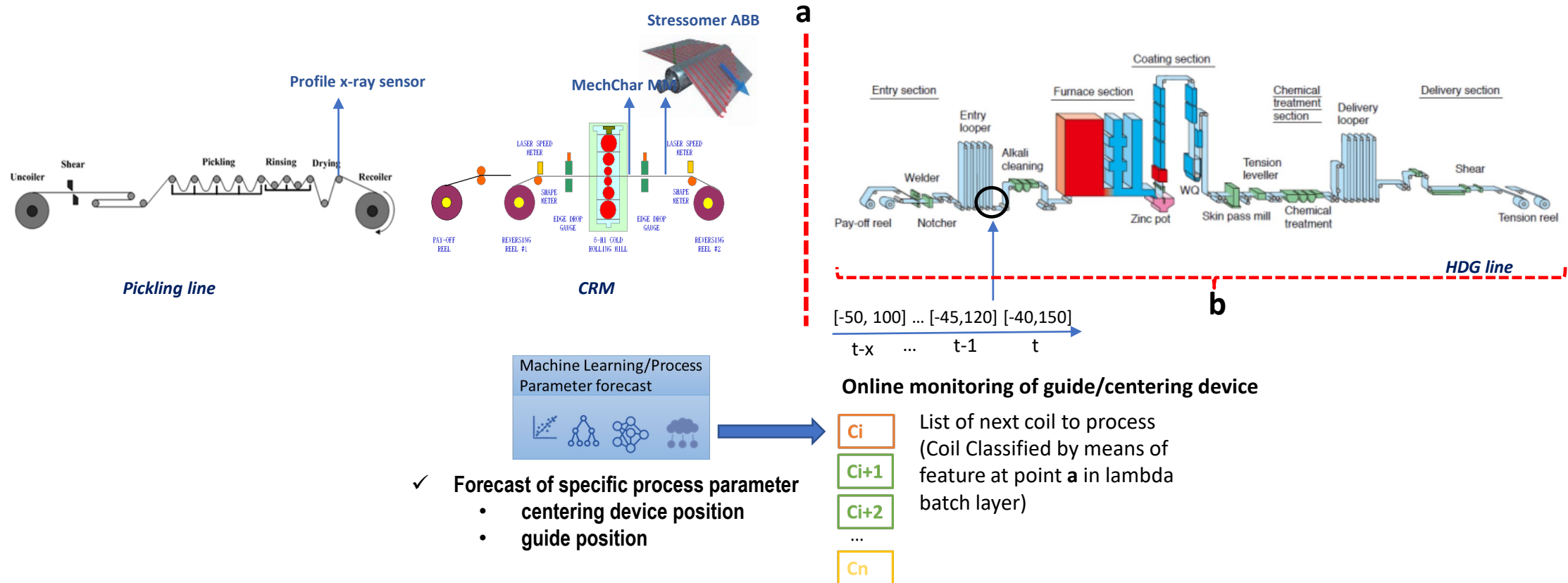Sensors and additional equipments position of the examined use case
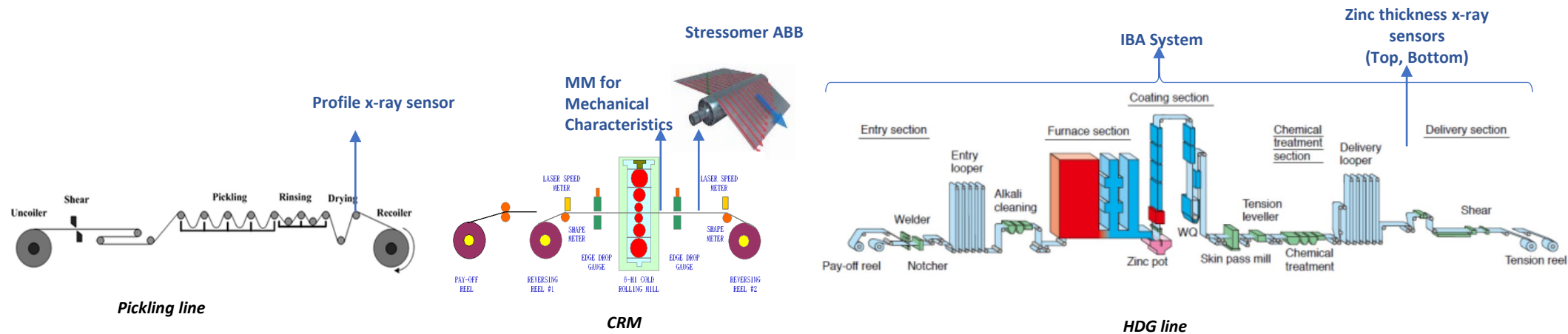
Identified Data Source

- ❖ **Pickling data**: Thickness along coil length (old system)
  - stored in files on server file system

- ❖ **Stressometer data at CRM for each pass**: Tension stress along coil length
  - stored in files on server file system

- ❖ **Mechanical characteristics**: Values along coil length
  - stored in files in proprietary format on server file system

- ❖ **Galvanization depth**: Zinc thickness on both sides
  - stored in files on server file system

- ❖ **Process parameters at HDG line from IBA System**: a set of process parameters
  - exported file from IBA QDR stored on server filesystem/streamed to Apache KafKa

- ❖ **Production Data**: Black coil data (dimension, mechanical characteristics...), production data, quality report, laboratory tests
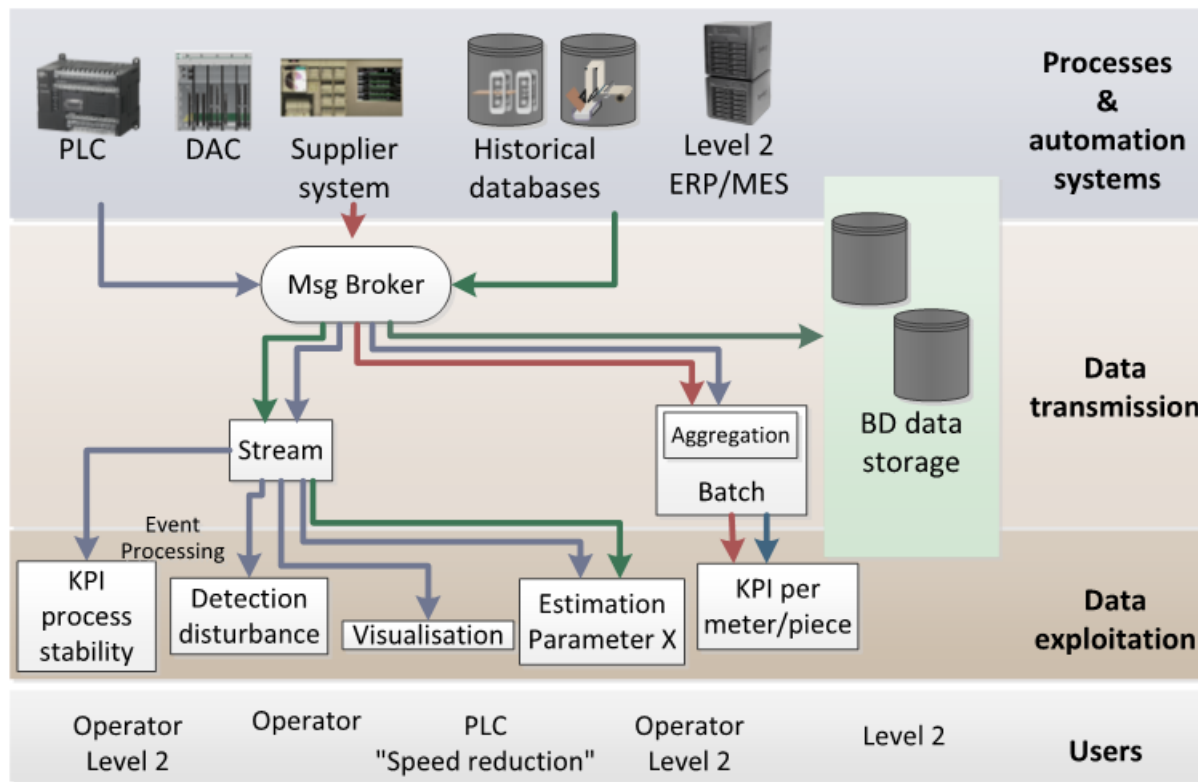 - per coil
  - stored on Sql database

Methodological approach

Development of a platform to strengthen and optimize the processes of coils galvanization

a. **Prediction of top defects at HDG line due to flatness problem**
b. **Prediction of specific parameters setup to avoid or limit problems during the galvanization**

Online monitoring of guide/centering device

Machine Learning/Process Parameter forecast

✓ **Forecast of specific process parameter**
- **centering device position**
- **guide position**

List of next coil to process (Coil Classified by means of feature at point **a** in lambda batch layer)

Ci
Ci+1
Ci+2
...
Cn

Suitable tools/platforms for BD framework

The main modules involves:

Collection/Ingestion: collect, preprocess, validate and format the raw data, stores into the Data Lake and sends to the broker
- Apache NIFI
- IBA to Kafka

Storage:
Data Lake: to gather all data coming from the ingestion layer into one system to perform data analysis and integration
- HDFS

Database: to store current calculated and historical data in one single place
- MySql

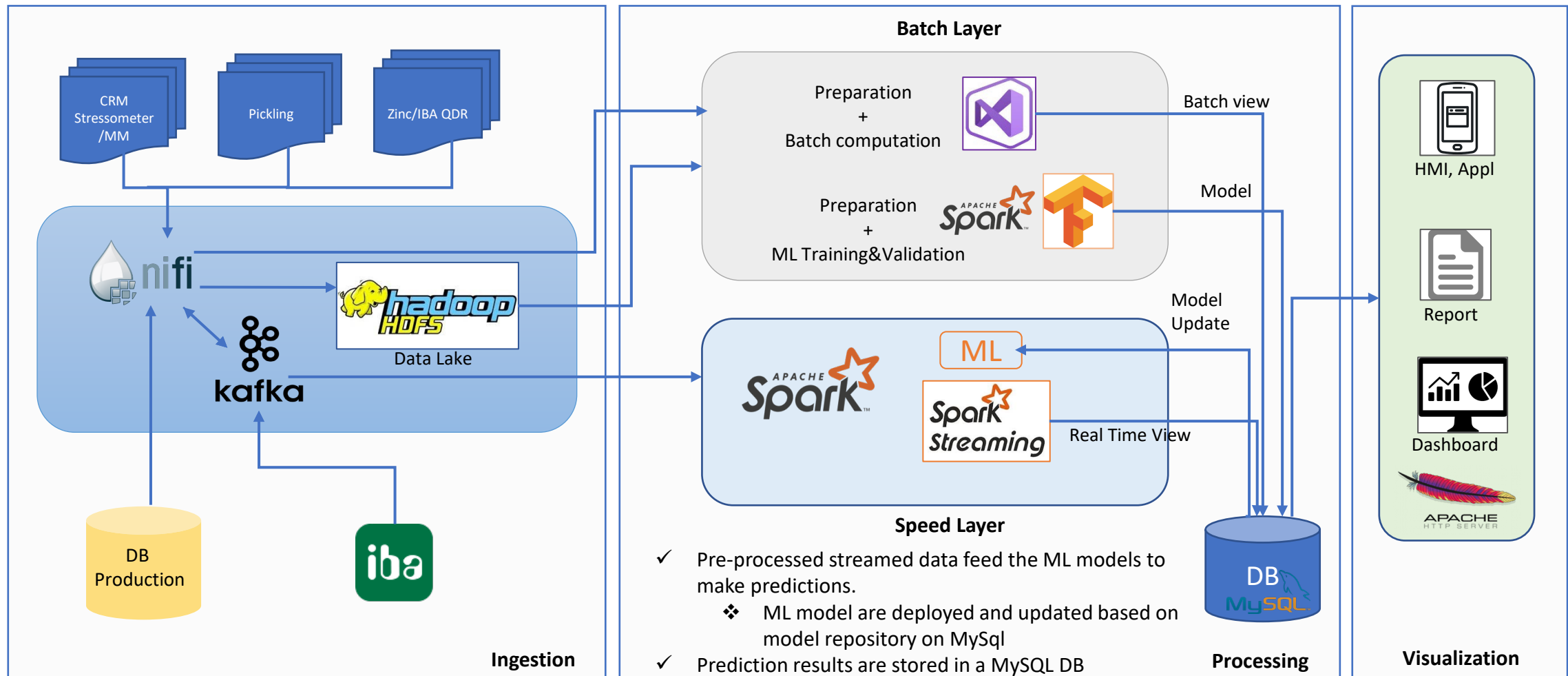**Analysis**:
Offline: Spark, C++, TensorFlow

Data Transmission: Kafka

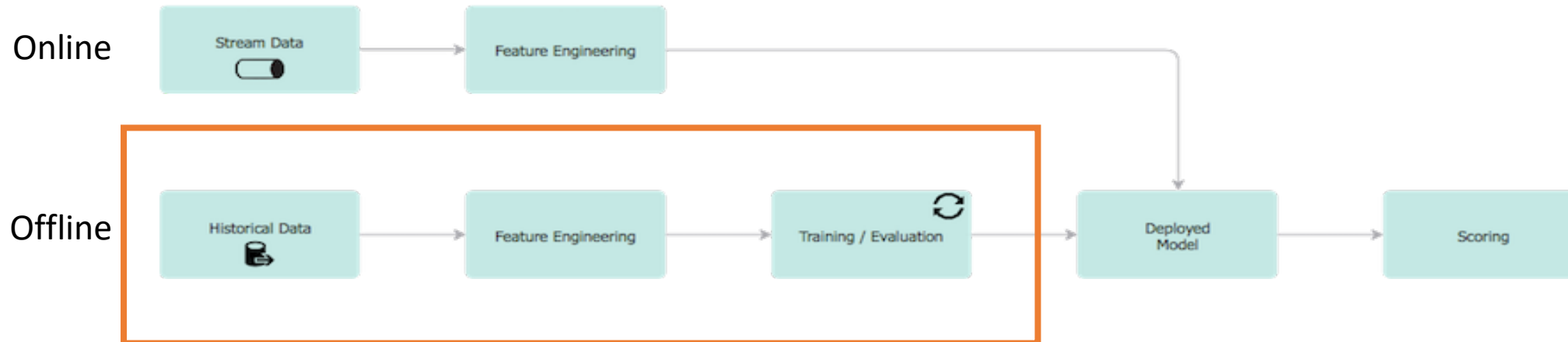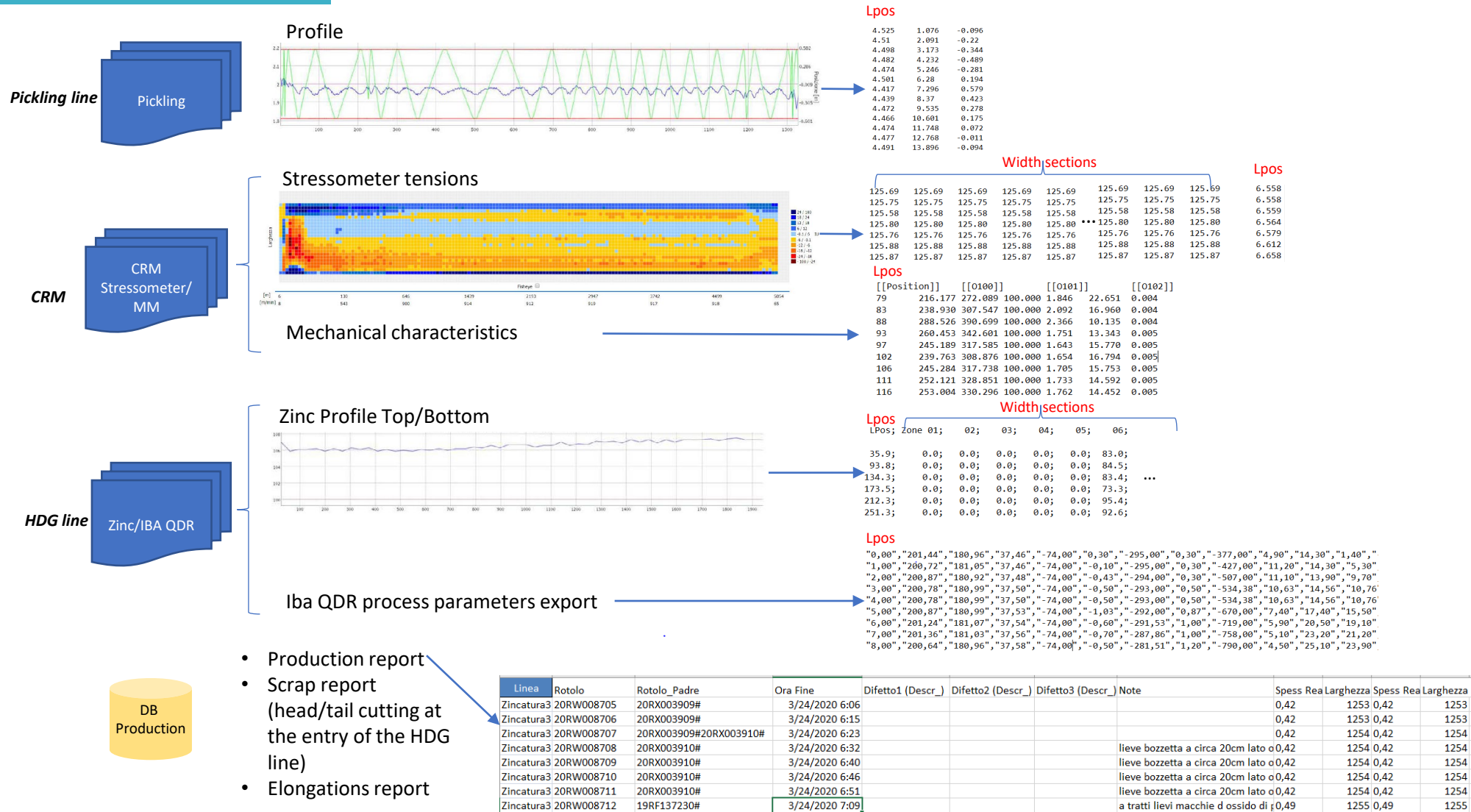**DataProcessing**: NIFI/C++(batch), Spark Streaming (realtime)

Data Visualization: php/HTML

Data Analytics for analysis and modelling

Online

Stream Data	→	Feature Engineering	

Offline

| Historical Data | → | Feature Engineering | → | Training / Evaluation | → | Deployed Model | → | Scoring |

Data preparation for analysis and modelling

Preparation of dataset based on historical data for the training/test of Machine Learning models involves:

- Rescale/Transform
  - Stressometer tension data are transformed to obtain IUnit values
- Data Alignment
  - Data on different sampling on length position are interpolated to obtain the maximum sample rate
  - Data sampled on different width section are interpolated based on the Stressometer sample rate
  - At HDG line coils are head/tail cut, a procedure realign the correct position with the IBA QDR and Zinc profile data based on two reported data:
    - the elongation at pickling line
    - the weight of the cut material at HDG line
- Missing values handling & Filter on noise
  - Row with missing values are dropped
- Outliers detection and treatment
  - CRM Stressometer external sections can contain suspicious values, they are excluded from the case
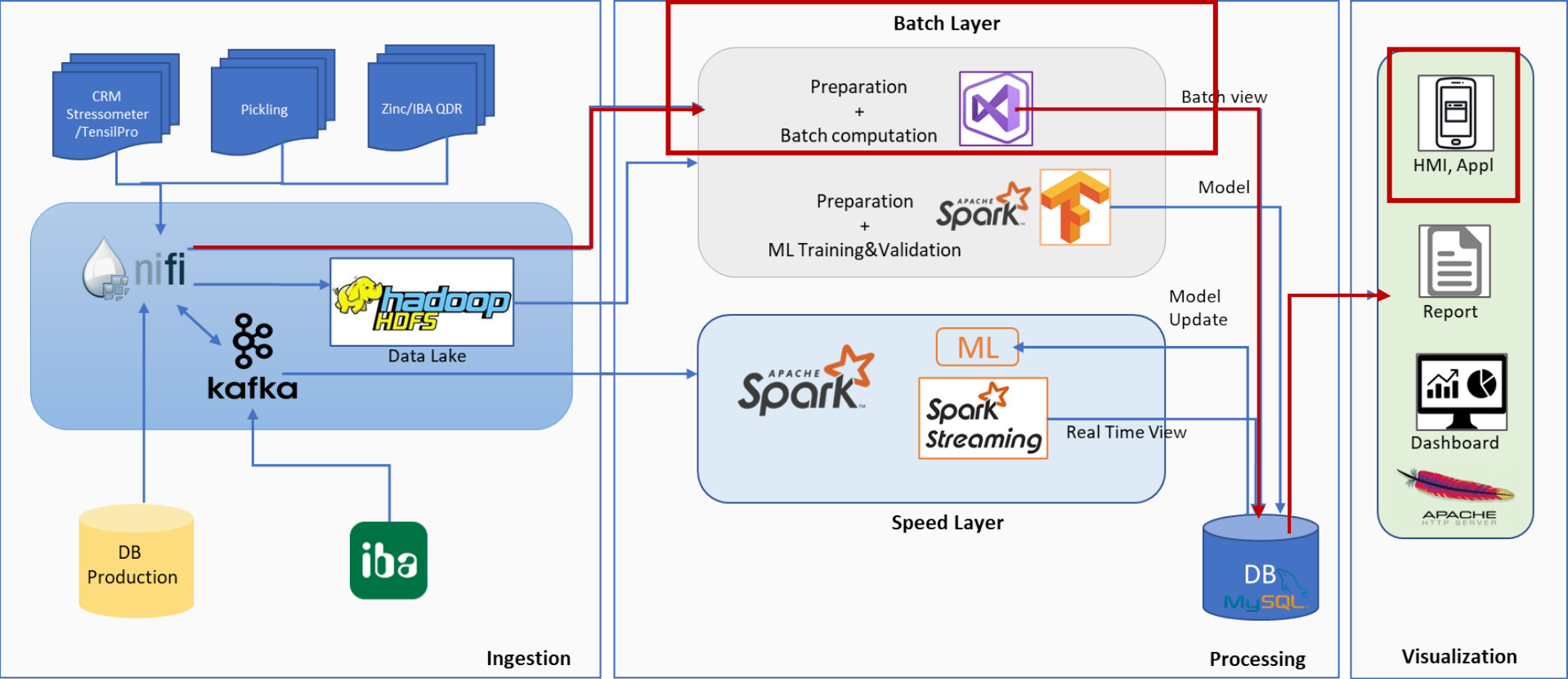
Analysis for feature/variable extraction:

- Feature engineering & Dimensionality reduction
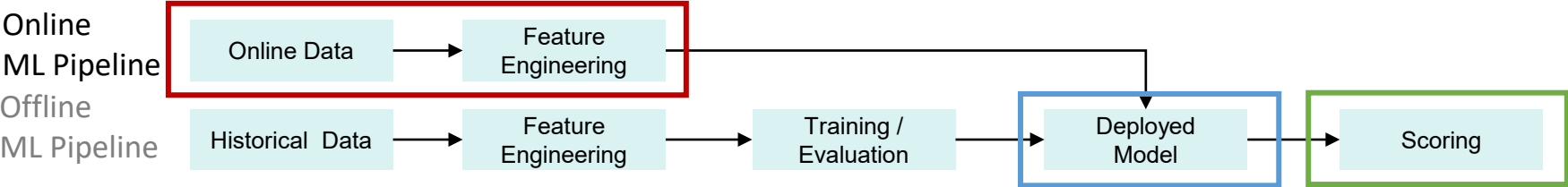 Correlations, PCA

Modelling

- Data Segregation
  - Split subsets of data to train the model and further validate how it performs against new data
- Training
  - Different schemes of Back Propagation Neural Network

Batch Layer Deploy

ML flow of batch process (a)

CONTROL IN STEEL

Online
ML Pipeline

| Online Data | → | Feature Engineering |

Offline
ML Pipeline

| Historical Data | → | Feature Engineering | → | Training / Evaluation | → | Deployed Model | → | Scoring |

CRM Stressometer/ MM

- **Rescale/Transform**
  - Stressometer tension to IUnit values
- **Data Alignment**
  - Interpolation on length/width

DB Production

→**Profile**
- Crown
- Wedge
- Build_up
- Thickness

→**Line, Coil_Qlt_In, Coil_Supplier**
- Vel_Process_Before_Acc Min/Max
- Vel_Process_HF_After_Acc_Min/Max
- Perc_Acc_Min/Max
- Vel_Process_Before_Acc_SET/Vel_Process_HF_After_Acc_SET/ Perc_Acc_SET

List of configuration for setup parameters

50_150_170
63_150_170
77_150_170
90_150_170
80_150_170

ANN0A

OUTPUT
- pos_guide1
- pos_centering1

ANN3

OUTPUT
- pos_guida3
- pos_centering3
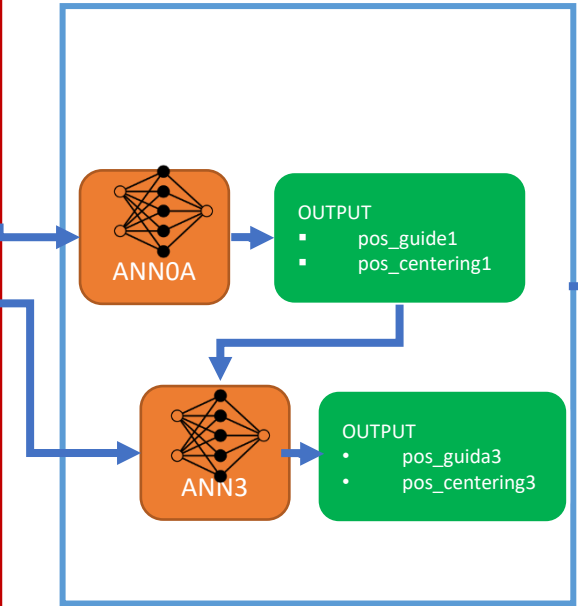
❖ Calculation of index on each position of coil length based on the value of guide and centering device predicted by ML model

Index	Position	Guide	Note
1	>=-20 <=20	>=-80 <=80	**Position:** Centered **Guide:** Low Correction
2	>=-20 <=20	>-200 <-80; >80 <200	**Position:** Centered **Guide:** Medium Correction
3	>=-50 <-20; >20 <=50	<=-80 >=80	**Position:** Slight Heel **Guide:** Low Correction
4	>=-50 <-20; >20 <=50	>-200 <-80; >80 <200	**Position:** Slight Heel **Guide:** Medium Correction
5	>=-180 <-50; >50 <=180	<=-200 >=200	**Position:** Severe Heel **Guide:** Max Correction

❖ Calculation of global coil criticality (red, yellow, green) based on the percentages of positional indexes

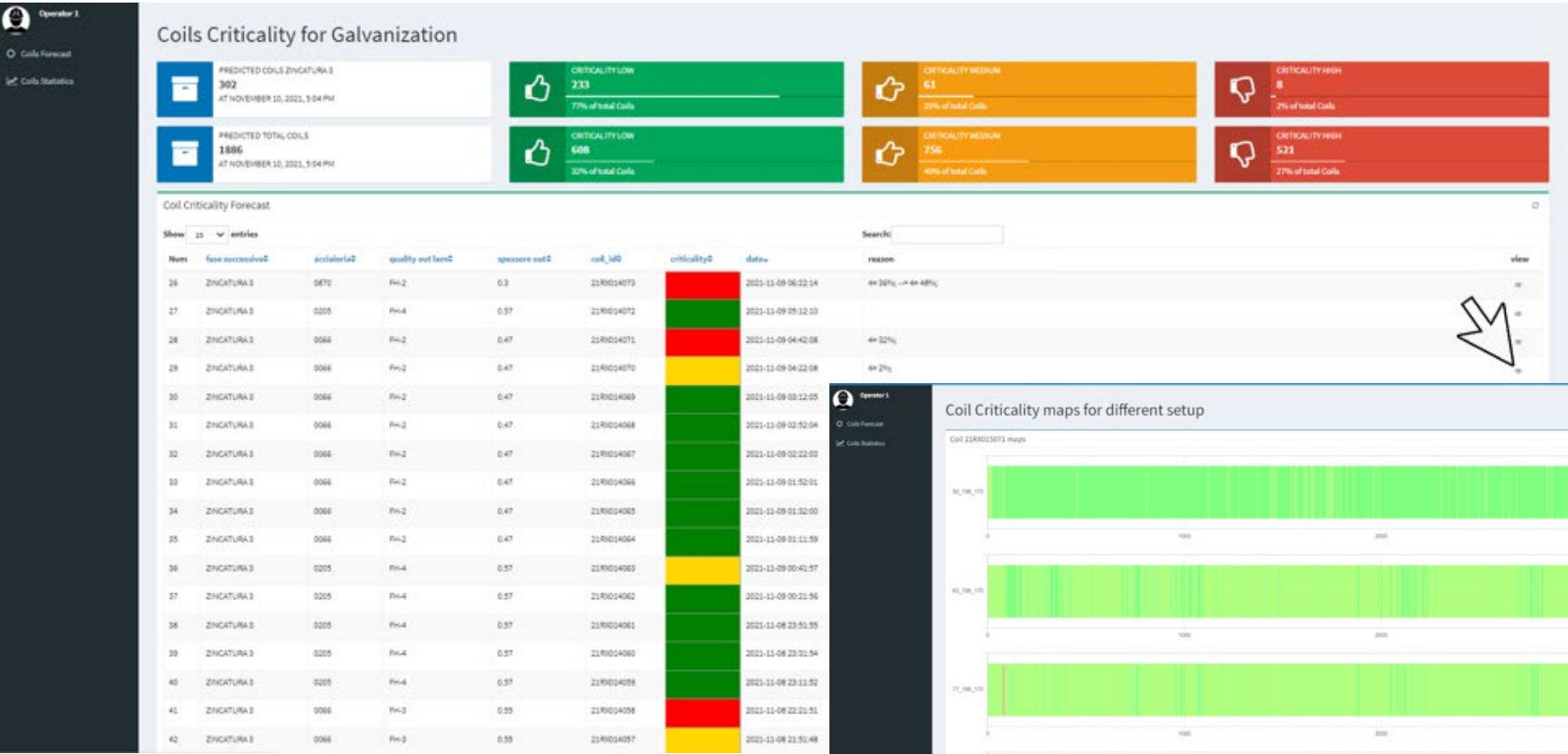| Criticality classification logic			
Index	% Coil Length		
1	0-100		
2	0-100		
3	80 -100	50 - 79.9	0-49.9
4	15 -100	2 -14.9	0-1.9
5	5 -100	1 - 4.9	0-0.9
0	0-100		

Flow implementation of batch process (a)

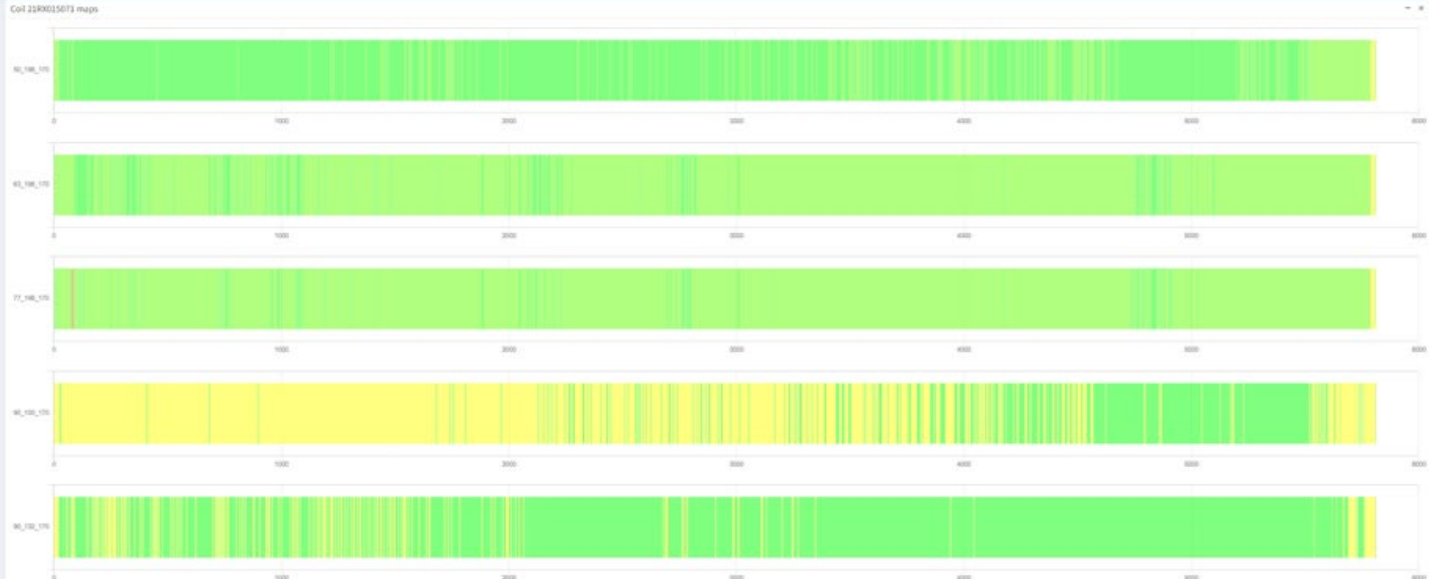Nifi execution flow with processors interactions starting from data sources until the ML model execution
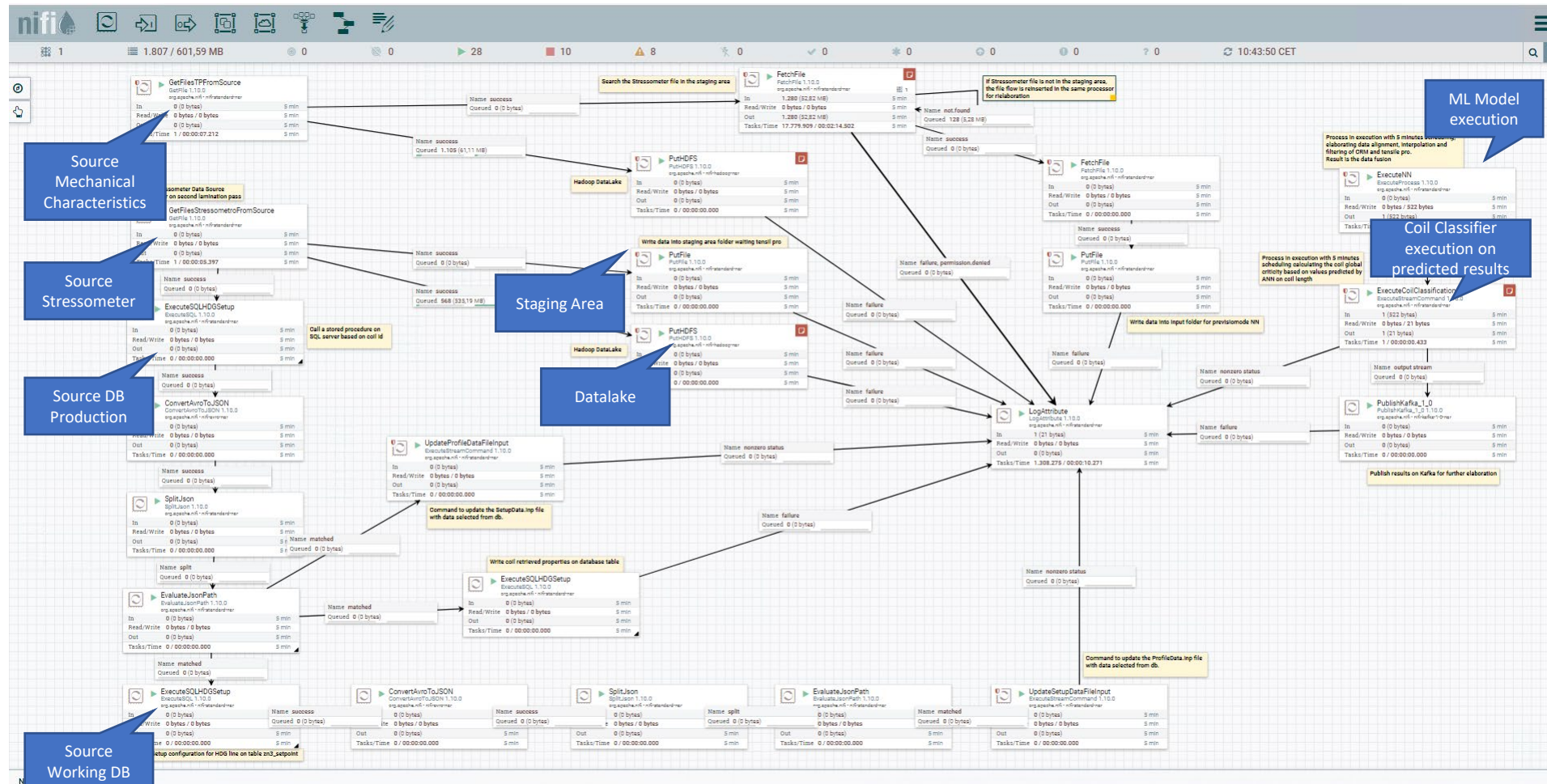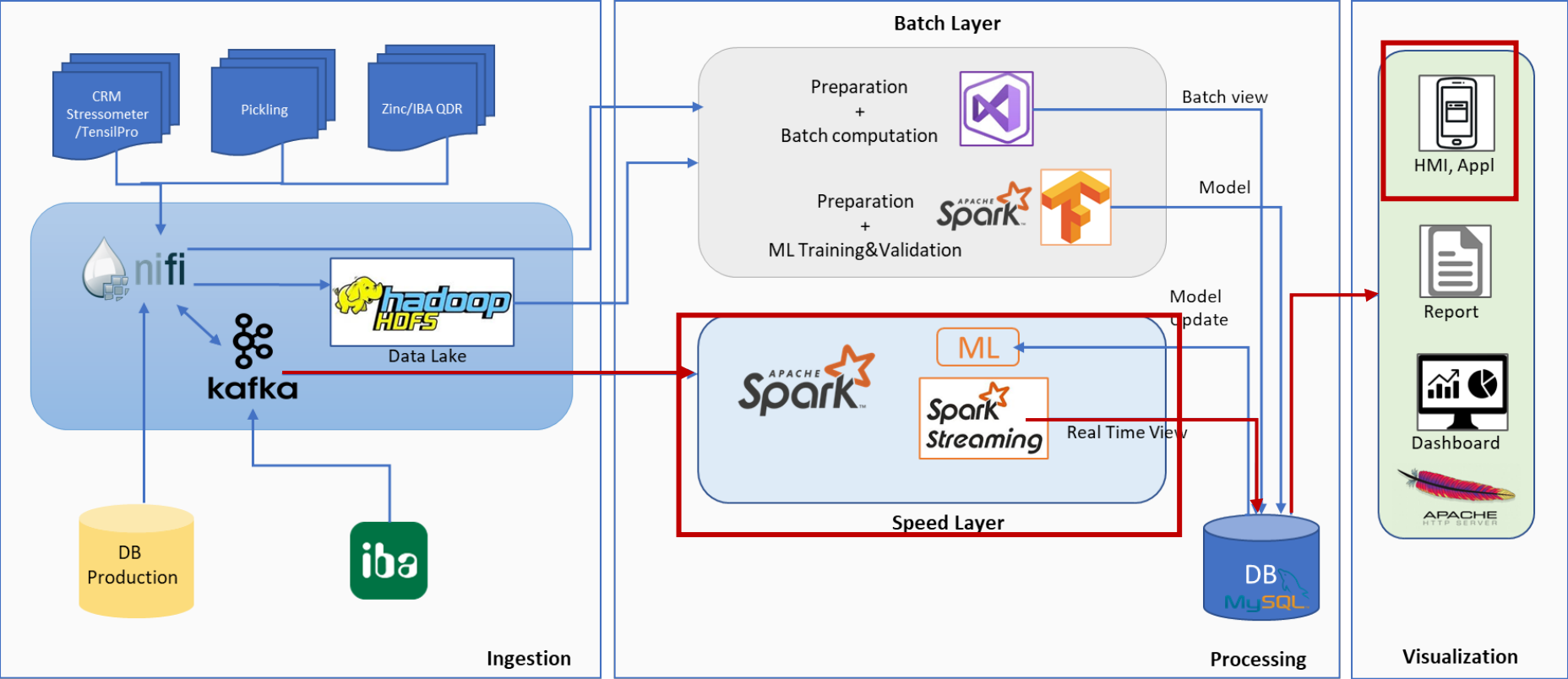
Visualization of batch process (a)

Web Interface based on Apache Http Server

Speed Layer Deploy

Implementation details of real time process (b)

❖ **Spark Structured Streaming** (introduced with Spark 2.x)
  - Model of streaming built on the Spark SQL library, based on Dataframe and Dataset APIs (Since Spark 2.0 Dataframe and Dataset can represent static, bounded data, as well as streaming, unbounded data)
  - Easily apply any SQL query
  - Spark SQL engine will take care of streaming, running in incrementally and continuously way and updating the final result as streaming data continues to arrive
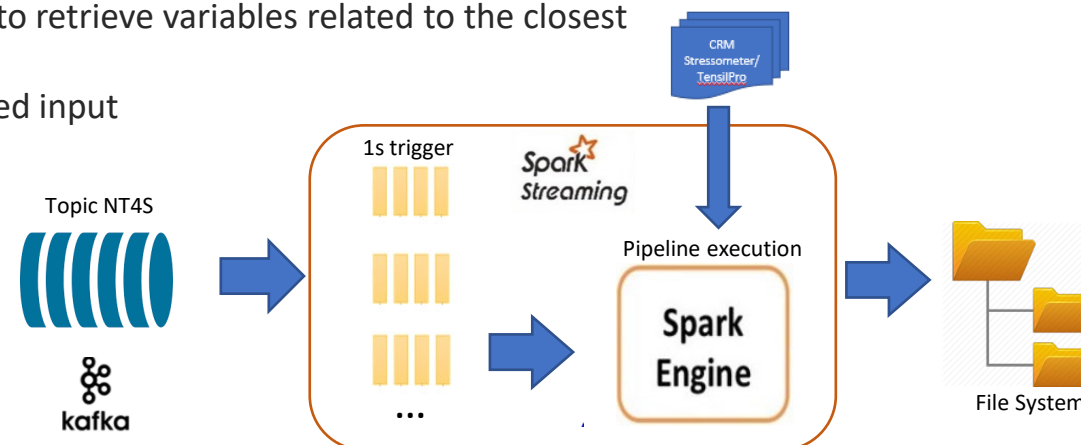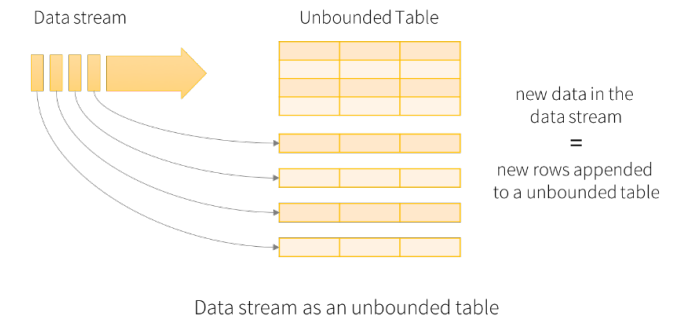
❖ **Using ForeachBatch** (Operations that allow you to apply arbitrary operations and writing logic on the output of micro-batch)
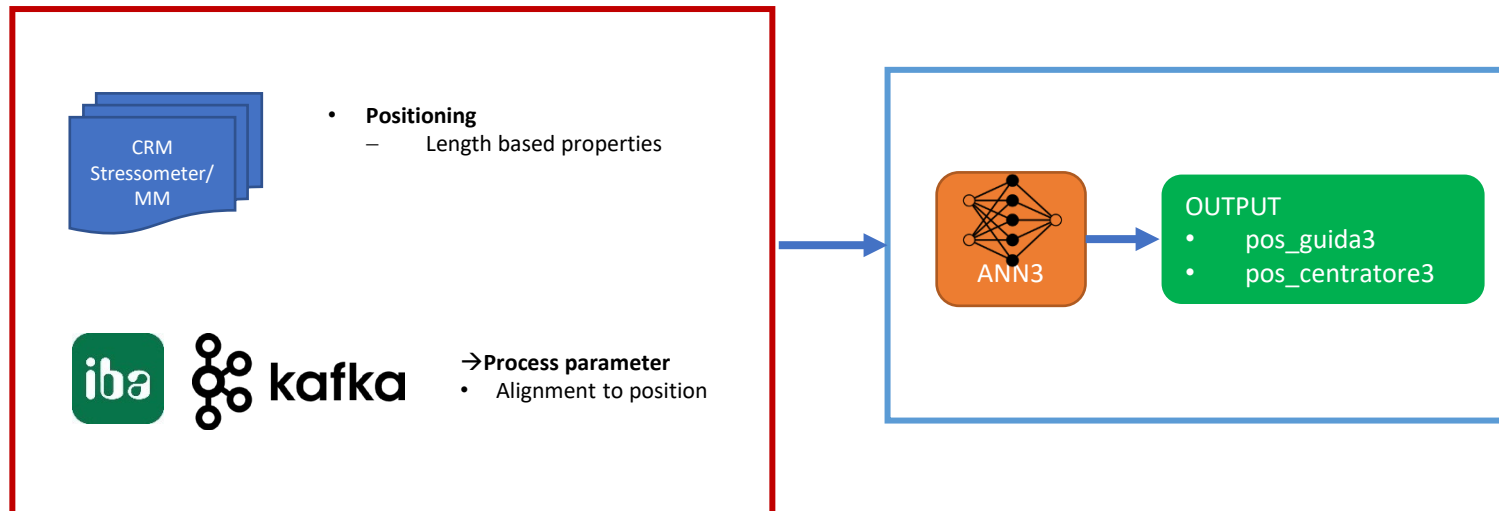 - ➢ **Pyspark ML pipeline**
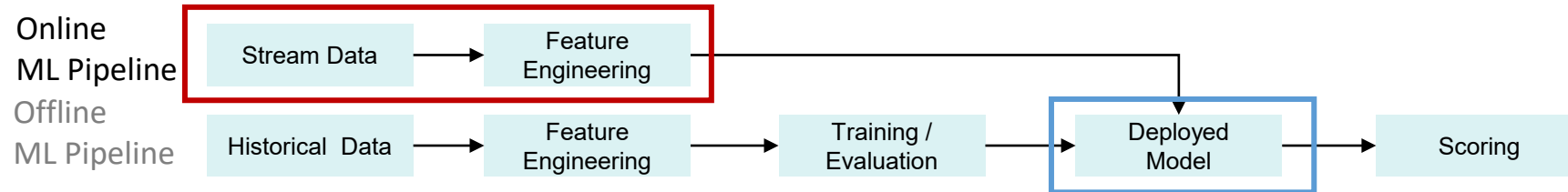   Foreach micro-batch received by topic
    - Get CoilId and entering length
    - Access Mechanical Characteristics, CRM file to retrieve variables related to the closest position to the entering length
    - Call the ANN procedure passing the composed input
    - Write results on file system

Structured Streaming Programming Guide - Spark 3.2.0 Documentation (apache.org)

Data stream as an unbounded table

ML Flow of real time process (b)

Online ML Pipeline

| Stream Data | → | Feature Engineering |

Offline ML Pipeline

| Historical Data | → | Feature Engineering | → | Training / Evaluation | → | Deployed Model | → | Scoring |

CRM Stressometer/ MM

- **Positioning**
  - Length based properties

iba kafka

→**Process parameter**
- Alignment to position

ANN3

OUTPUT
- pos_guida3
- pos_centratore3

Summary of results

- ✓ Survey about necessary/available data was generated including information about location, origin, content and amount
- ✓ Definition of requirements about integration of new tools and architectures into the existing plant IT systems.
- ✓ Definition of HMI under different aspects (security, types of users, visualizations etc)
- ✓ Benchmarks and selection of tools and platforms for data storage, transfer and analysis
- ✓ Setup of IT system with necessary measures (additional computation server, expanded storage, elongation of historical data retention, proprietary acquisition system)
- ✓ Installation and tests for assessing tools and platforms for data handling, storage and analysis
- ✓ Development of auxiliary tools for the integration and **exploitation** of data into the new system
- ✓ Deployment of implemented features
- ❖ Validation of results

CONTROL IN STEEL

Thank you for
your attention!

RINA Consulting - CSM S.p.A.
Via Di Castel Romano 100 - 00128 ROMA - ITALIA

www.rina.org

RINA. Excellence
Behind Excellence.